perm.pub/D9qSdCY6GPrxthT3ZnFouEU35ow/1.1
Additional formats and editions available online.

Edition 1.1

Author date: 2023-09-29

Lineal admixture time: an interdisciplinary definition

E. Castedo Ellerman (castedo@castedo.com)

Citation:
E. Castedo Ellerman (2023) "Lineal admixture time: an interdisciplinary definition" https://perm.pub/D9qSdCY6GPrxthT3ZnFouEU35ow/1.1

Copyright:
© 2023, Ellerman et al
CC BY License
This document is distributed under a Creative Commons Attribution 4.0 International license.

Abstract

STAGE: DISCUSSION DOCUMENT

OBJECTIVES:

- Define *lineal admixture time*.
- Invite discussion and feedback, especially regarding questions at the end of this document.
- Facilitate the discovery of mathematical tools, using this definition, that are useful to downstream empirical research.

SIGNIFICANCE: This definition is interpretable within empirical research without assuming particular mathematical models of demographic or genetic processes.

DOWNSTREAM RESEARCH: The estimation of the timing and extent of interbreeding between previously separated populations.

Background

The genealogies of a population reflect population events of the past. Lineages, at the organismal level, are lines of descent carrying genetic information through a population's genealogies. These lineages are a link between genetic evidence and population events of the past. In contrast to estimates of purely genetic quantities, estimates of quantities about lineages can have closer correspondence to non-genetic evidence, such as archaeological, geological, and cultural evidence.

Definition of lineal admixture time

We define *lineal admixture time* per lineage. This measure in time duration depends on two reference time points:

1. a past time horizon when all ancestors are members of separate non-admixed subpopulations, and
2. an observation time, such as the present.

Given a lineage, the **lineal admixture time** is defined as:

the amount of time since fertilization of the first admixed individual in the lineage (or zero if the lineage has no such individual).

Stated more compactly, lineal admixture time is time since a lineage's first admixed fertilization. Admixed individuals are all individual whose parents are not from the same non-admixed subpopulations.

The *average lineal admixture time* of an individual is an average over all lineages ending at that individual. At every merging of lineages through offspring, equal weighting is given to the respective sets of maternal and paternal lineages. The *average lineal admixture time* of a group of individuals is the average lineal admixture time across all those individuals.

It is worth noting that non-zero admixture times do not remain constant if the observation time changes. It follows that the non-zero average admixture time of an individual is not fixed. It will increase if the observation time increases with age.

An example with Mendel's peas

We consider an example of lineal admixture times with Mendel's famous peas [1]. We imagine one of his classic experiments starting in 1860 and ending in 1862. Because the pea plant is an annual plant, each generation is separated by one year. During the lifetime of one generation, fertilization of the next generation occurs, in the form of seeds (illustrated with a `.` symbol).

```
1860  A-.-Y   B-.-Z
        |       |
1861    M---.---N
            |
1862        Q
```

In this example we will treat the non-admixed ancestral subpopulations to be true-breeding pea plant varieties in 1860. Plants `A` and `B` in the diagram are non-admixed plants with round peas. Plants `Y` and `Z` are non-admixed plants with wrinkled peas. These two subpopulations grow and are crossed in 1860 producing seeds that same year. Because they were crossed, the offspring `M` and `N` are both admixed (hybrid). This first admixed (hybrid) generation will produce seeds in 1861 for the second admixed generation, plant `Q`.

Plant `Q` has four lineages:

- `Q-M-A`
- `Q-M-Y`
- `Q-N-B`
- `Q-N-Z`

In these lineages the first admixed individuals are `M` and `N`, both of which were fertilized in 1860. The first admixed fertilizations occur in 1860. Thus the lineal admixture time for all lineages of `Q`, observed in 1862, is **2 years**.

Hybrid generation numbers

A related definition measures generation number rather than time. Given a lineage, the *lineal hybrid generation number* is defined as:

the number of admixed individual in the lineage.

Because peas are annual, the lineal admixture times in years coincide exactly with the lineal hybrid generation numbers. The first and second hybrid generations are often referred to as F_1 and F_2 [2]. Lineal average hybird generation numbers are a generalization of indexes given to hybrid generations F_1, F_2, etc...

An example with humans

These definitions are inspired by the use of generation numbers to infer admixture timing in the study of admixture in Greenland [3].

We give a hypothetical example in Greenland at an observation time of 1600 CE. We categorize the ancestors at a time horizon of 1000 CE into separate non-interbreeding Inuit and European subpopulations. We denote the fertilization and death of an individual with symbols `.` and x, respectively, and write names at the time an individual is a 3-month-old baby.

```
1500  Erik
1510   |       Aaju
1520   |         |   Natar     Atuat
1530   |----.----|     |         |
1540   |  Tagak  |     |----.----|
1550   |    |    x     x  Mikak  |
1560   x    |               |    |
1570        |-------.-------|    x
1580        x     Kiviaq    |
1590                |       x
1600                |
```

We imagine Erik as only having European ancestors and Aaju, Natar and Atuat only having Intuit ancestors. As of 1600 CE, the lineages are:

- Kiviaq-Tagak-Erik
- Kiviaq-Tagak-Aaju
- Kiviaq-Mikak-Natar
- Kiviaq-Mikak-Atuat

For the Tagak lineages, the first admixed fertilization is in 1530. In contrast, the Mikak lineages have a first admixed fertilization in 1570. Thus the lineal admixture times are equal parts 70 years and 30 years, for an average lineal admixture time of 50 years. The respective lineal hybrid generation numbers are 1 and 2 resulting in an average lineal hybrid generation number of 1.5.

Advantages of lineal admixture time

Lineal admixture time has a number of advantages as a quantity for research.

Firstly, it is independent of any particular model of an admixture process. With enough genealogical information one could calculate lineal admixture times of a real group of individuals. There is one true set of lineal admixture times for a real population. But realistically, we can only hope to estimate those numbers based on evidence and models. But the definition exists independent of any particular model used for estimation.

Secondly, the interpretation of lineal admixture time does not require fluency in probability theory. Researchers with interests in empirical evidence and not mathematics can make use of estimates of lineal admixture times.

Thirdly, we conjecture that *distributions* of lineal admixture times across individuals, and groups of individuals will prove to be a useful mathematical tool in the timing of admixture. This conjecture is based on not-yet-documented mathematical work by ECE. Lineal admixture time is conveniently representable as a random variable from a stochastic process in which lineages are random objects.

The same population events that affect admixture also affect hybridization, introgression, and gene flow. We conjecture that statistical tools for estimating distributions of lineal admixture time will also be useful to these related topics in addition to admixture timing.

The Greenland example illustrates how average lineal admixture time is easily testable against non-genetic lines of evidence. If a genetic model estimates the average lineal admixture time of present day Greenlanders to be 500 years, we can consider historical and archaeological evidence to falsify that estimate.

End

This document presents a definition of *lineal admixture time*. This quantity is the basis for current mathematical studies of ECE. Feedback and input is greatly appreciated, especially regarding the relevance to downstream empirical research. In particular, the following questions are of interest:

1. Is *lineal admixture time* easily understood and interpretable without fluency in probability theory?
2. Does *lineal admixture time* benefit from being comparable to non-genetic lines of evidence (i.e archaeological, geological, linguistic, historical, cultural)?
3. What are existing terms for this specific definition of admixture time?
4. Are the following terms:
    - lineage
    - genealogy
    - lineal admixture time
    - lineal hybrid generation number

   potentially confusing in the way they are used in this document?

References

1. Abbott, Scott, and Daniel J. Fairbanks. “Experiments on Plant Hybrids by Gregor Mendel”. *Genetics*, vol. 204, no. 2, Oct. 2016, pp. 407–422, https://doi.org/10.1534/genetics.116.195198.
2. Hartl, Daniel L., and Elizabeth W. Jones. *Essential genetics: A genomics perspective*. 4th ed., Boston: Jones; Bartlett Publishers, 2006.
3. Waples, Ryan K., Aviaja L. Hauptmann, Inge Seiding, Emil Jørsboe, Marit E. Jørgensen, Niels Grarup, Mette K. Andersen, Christina V. L. Larsen, Peter Bjerregaard, Garrett Hellenthal, Torben Hansen, Anders Albrechtsen, and Ida Moltke. “The genetic history of Greenlandic-European contact”. *Current biology: CB*, vol. 31, no. 10, May 2021, pp. 2214–2219.e4, https://doi.org/10.1016/j.cub.2021.02.041.